Against the Grain

Volume 18 | Issue 2 Article 40

April 2006

Standards -- Libraries, Data Providers, and SUSHI: the Standardized Usage Statistics Harvesting Initiative

Adam Chandler
Cornell University Library, alc28@cornell.edu

Tim Jewell
University of Washington Libraries, tjewell@washington.edu

International Dateline
from page 81

lic domain

- Use the extended **RSS** standard to provide a greater level of automation of interoperability between the publisher and a library through the integrated library system
- Deliver open source software components and **XML** schemas, which will be freely available for the whole of the academic library, publisher and integrated library system supplier communities
- Produce a demonstration implementation of a service utilising the standards and software as part of the project
- Report of the impact of this project for a particular academic institution

Dictate: Distributed Content Tagging Tool for EPrints — Social bookmarking systems such as **Connotea** and **del.icio.us** allow users to organize, share and discover Web resources by using "tags" (i.e., keywords) to label and share them. This approach has, in turn, lead to the concept of a "folksonomy" — the collection tags associated by a group of users with a set of content, and the relationship between them. As a way of organising information, this is distinct from, and perhaps complementary to, traditional top-down, hierarchical taxonomies. **GNU EPrints** is a popular system for creating institutional repositories. However, in common with most institutional repository software, it has yet to employ tagging to enable users to organize better and discover the content held there.

The object of this project is to add bookmarking, tagging and recommendation features to institutional repositories running **GNU EPrints**, by extending that software to enable it to integrate with the existing social bookmarking systems. This should enable more effective organization and discovery of content held in **EPrints** institutional repositories.

Something Useful to Read on the Train

Gwendoline never travelled without her diary because "One should always have something sensational to read on the train." For those of us who are more interested in useful rather than sensational reading, I recommend leaving the diary at home and instead taking advantage of wireless connections to catch up on the very practical research being sponsored by **JISC**. Further information on the projects described above, as well as the other metadata and interoperability projects currently supported by **JISC** can be found on the **JISC** Website.[3] All are due to report their findings by 31 July 2006; some have already completed their work and the resulting tools are being applied in practice.

References

1. Metadata Demystified: A Guide for Publishers; Brand, Daly and Meyers; NISO/The Sheridan Press (2003).
2. *http://www.jisc.ac.uk/about_info_env.html*
3. *http://www.jisc.ac.uk/index.cfm?name=programme_pals2*

Standards — Libraries, Data Providers, and SUSHI: the Standardized Usage Statistics Harvesting Initiative

by **Adam Chandler** (Information Technology Librarian, Library Technical Services, Cornell University Library, 107B Olin Library, Ithaca, NY 14853-5301; Phone: 607-255-5760) <alc28@cornell.edu>

and **Tim Jewell** (Director, Information Resources, Collections and Scholarly Communication, University of Washington Libraries, Box 352900, Seattle, WA 981195-2900; Phone: 206-543-3890) <tjewell@u.washington.edu>

The sheer amount of usage data now available to librarians who manage electronic resources represents a true "embarrassment of riches:" while the information is extensive and necessary to their operations, keeping up with it all has become a major problem. E-resource management systems now available or in development will provide tools to store and manipulate the data but may not offer much help with "intake."

The **Standardized Usage Statistics Harvesting Initiative (SUSHI)** discussed below is aimed at overcoming this problem, and while librarians and ILS vendors have already demonstrated strong interest in it, there are reasons why data providers like e-journal and database publishers and vendors should also embrace it. For example, as awareness and interest in **SUSHI** build within the library community, data providers that become early supporters and adopters can demonstrate their vision and responsiveness to the market, while those that don't may be rightly perceived as lagging behind. And while there will be development costs associated with supporting **SUSHI**, those costs should be minimal and may lead over time to the need to invest less in developing custom usage data interfaces. Lastly, since the **SUSHI** protocol has already garnered support by **COUNTER**, it could conceivably become a requirement for **COUNTER** certification; work done now to implement it will position vendors and publishers to retain or establish certification.

COUNTER: Opportunities and Challenges

Under the skillful leadership of **Peter Shepherd** and the **COUNTER Board**, **COUNTER** is securely fixed among publishers and librarians as the standard for reporting online usage.[1] However, on the demand side of the fence, there are serious impediments to making use of the **COUNTER** reports. The formatting of the delimited reports (either comma separated or in **Excel**) is widely variable across publishers, which makes processing them automatically difficult or impossible. Though steps are being taken to make the reports more consistent, **COUNTER** reports are not yet available in a consistent data "container." The administrative cost of individual provider-by-provider downloads is also high. To make complete use of the reports, library staff must log into each and every one of their different providers, navigate to the right page, then download the appropriate report to their hard drive. Then it is up to the library staff member to consolidate the individual report with all the others and for all the other months.

The **Standardized Usage Statistics Harvesting Initiative ("SUSHI")** represents a "Web services" approach to solving the **COUNTER** consolidation problem.

SUSHI as COUNTER Companion

In response to the difficulty librarians are experiencing in making use of the **COUNTER** delimited reports, last summer a small group of librarians and vendors assembled and sketched out an automated protocol for moving **COUNTER** XML reports from providers

continued on page 83

Libraries, Data Providers, and SUSHI: *from page 82*

to libraries.[2] Later the group agreed to call the effort the **Standardized Usage Statistics Harvesting Initiative (SUSHI)**. After a promising start, **NISO** recognized the activity in November.[3]

The protocol is essentially a **SOAP (Simple Object Access Protocol)** request/response Web Services "wrapper" for the XML version of **COUNTER** reports. In the protocol a transaction begins when a client service such as a usage data consolidation service or ILS vendor identifies itself to a data provider, identifies the customer whose statistics are being asked for, and specifies the desired report. In response, the server provides the report in XML format, along with the requestor and customer information — or an appropriate error message. We envision a system in which the client system is programmed to automatically retrieve last month's report, for all the **COUNTER** compliant vendors with which the library does business.

While there are a variety of possible requesters, the **SUSHI** model is premised on the notion that a long term solution mandates integrating usage statistics into existing library workflows, and might best be handled within the framework of an **Electronic Resource Management (ERM) System**. Given this background, it should come as no surprise that the librarians on the **SUSHI** working group are all steering group members of the **Digital Library Federation's Electronic Resource Management Initiative**. In fact, **SUSHI** is an integral part of phase 2 of the **DLF ERMI**.[4] Librarians are seeking a comprehensive solution to the management of licensed electronic resources that combines licensing, accurate holdings, orders, and statistics, among other important information from their entire life cycles. We believe models that set statistics apart from the rest of the life cycle of electronic resources are substantially less valuable, since so many factors must be considered when evaluating them.

Status of SUSHI

There are five vendors actively engaged in developing the **SUSHI** protocol: two content providers (**EBSCO Information Services** and **Swets Information Services**), two ILS vendors that offer **ERM** products (**Innovative Interfaces, Inc.** and **Ex Libris**), plus **Thomson Scientific**, which is working on a **COUNTER** statistics component that would link to their journal impact index. The initial emphasis among the development partners is on transmission of the **COUNTER JR1** report.

As we write, the protocol will be offered in production within weeks at **EBSCO** and **Swets**. **III** is making it available in the beta version of their **ERM** module this month. **Ex Libris** will be following later this year with its integration of **SUSHI** in their **Verde ERM** product. Other vendors committed to implementing the protocol in 2006 include **HARRASSOWITZ, Endeavor Information Systems**, **SirsiDynix**, **OCLC**, and **Serials Solutions**. The **Florida Center for Library Automation** and **College Center for Library Automation (CCLA)** from the **State of Florida Community Colleges** are also interested.

The challenge at this point is lining up support among content providers. Beyond **EBSCO** and **Swets**, **Cornell's Project Euclid** is the next content provider on the list to offer the protocol to customers. Early adopters of the protocol stand to gain a competitive edge, since librarian colleagues we speak with are enthusiastic about the potential of the **SUSHI** protocol and are sure to compare it favorably when making decisions about products for purchase. **SUSHI** development kits are freely available for the **.NET** and **Java Web Services** environments. More information about **SUSHI** is available on the project page.[5]

Endnotes

1. COUNTER Website: *http://www.projectcounter.org/articles.html.*
2. In the summer of 2005, the project which would come to be called **SUSHI** started its investigations. Members of the team included: **Ivy Anderson** (California Digital Library), **Adam Chandler** (Cornell University Library), **Ted Fons** (Innovative Interfaces Inc.), **Bill Hoffman** (Swets Information Services), **Tim Jewell** (University of Washington Libraries), **Ted Koppel** (Ex Libris), and **Oliver Pesch** (EBSCO Information Systems).
3. "NISO Initiative to Standardize Online Usage Statistics Harvesting." Press Release. *http://www.niso.org/news/releases/pr-stats-11-05.html*
4. DLF Electronic Resource Management Initiative, Phase II. *http://www.diglib.org/standards/dlf-erm05.htm*
5. Standardized Usage Statistics Harvesting Initiative (**SUSHI**). *http://www.niso.org/committees/SUSHI/SUSHI_comm.html*

IMHBCO (In My Humble But Correct Opinion) — Thinking About the Value of Staff Time

by **Rick Anderson** (Director of Resource Acquisition, University of Nevada, Reno Libraries; Phone: 775-784-6500 x.273) <rickand@unr.edu>

Those of us who work in acquisitions, serials, or electronic-resource departments are generally pretty good at thinking about the value of our materials budgets. With funding getting tighter and prices rising dramatically every year, we've had no choice but to make hard decisions about which products and services are most important to our users, and many of us have come up with effective (even ingenious) ways of determining which of those products and services will give our patrons the most bang for our budget buck.

But we have another limited resource that has to be managed carefully, and that's staff time. While it's true that prices are rising at a ridiculous rate, it's also true (thanks to the advent of "free-online-with-print" subscriptions and Big Deals) that most of us are providing quite a bit more content to our patrons than we did in the print era. When serials inflation outpaces growth in materials budgets, the result is that those budgets become more "scarce" even if they remain technically the same — we are able to buy less and less with the same amount of money every year. But a similar thing happens to staff time when the amount of content we offer rises while staffing levels remain the same, or when a given unit of content moves from print to online format, requiring more staff time and oversight. An hour of staff time no longer "buys" as much resource management as it did when we had fewer resources to manage, and when the vast majority of our resources were published in relatively low-maintenance print formats.

I alluded to this problem in an earlier column ("No, You May Not Come Train My Staff," vol. 16 #4, September 2004) in which I mentioned that I often decline sales reps' offers to come in and talk to my staff about new and upcoming products. The problem, I said, is not that we don't want to know about their products, but that staff time in our libraries is often severely limited and we may not be able to afford to invest twenty or thirty scarce staff hours in gathering people to listen to a sales pitch. Instead, we need to get product information in written form, so that it can be distributed to everyone and read when they have time.

But this begs a question: it's all well and good to say that staff time is scarce and valuable, but how do we determine what an hour of staff time is actually worth?

There are many different ways of answering this question. I'm going to propose two:

continued on page 84